Health Continuum and Data Exchange in Belgium and in the Netherlands
Francis H. Roger France et al. (Eds.)
IOS Press, 2004

Incorporating Evaluation into the Design of a Decision-Support System

Stefan VISSCHER, Karin SCHURINK and Marc BONTEN
Department of Internal Medicine and Infectious Diseases
University Medical Centre Utrecht
{S.Visscher, K.Schurink, M.J.M.Bonten}@azu.nl

Peter LUCAS
Institute for Computing and Information Sciences
Radboud University Nijmegen
peterl@cs.ru.nl

Jeroen VAN WOLFFELAAR and Peter VAN DE WERKEN
Institute of Information and Computing Sciences
Utrecht University
{jjawolff, pwerken}@cs.uu.nl

Abstract. Medical decision support systems will only be accepted by the medical community if properly evaluated. However, little attention has been given in the scientific literature to the topic of how to incorporate evaluation issues into the design of a decision-support system. In this paper, we describe work in developing a decision-support system that is intended to support the management (diagnosis and treatment selection) of ventilator-associated pneumonia in patients. From the beginning of the development of this system, we have taken care to incorporate evaluation issues into the design of the system. In the paper, we analyse the problems that need be taken into account when evaluating a system. Next, we describe the consequences for the functionality of the system.

Keywords. Decision-Support System, Decision Theory, Bayesian Network, Evaluation Biases, Ventilator-Associated Pneumonia, Antibiotics, Intensive Care Unit

Introduction

The management, i.e. the diagnosis and the selection of optimal treatment, of disorders in critical care is a challenging task, as patients are usually severely ill and often have a number of concomitant disorders; thus if treatment is delayed this may be the cause of death of a patient. Diagnosing a disorder is in particular difficult if there are few signs and symptoms that are typical for the disorder and if a disorder does not occur very frequently. Furthermore, selecting optimal treatment is difficult as there is normally no time to wait with instilling treatment until the results of laboratory tests become available. This is, for example, the situation with ventilator-associated pneumonia, or VAP for short. VAP is a form of pneumonia that occurs in patients whom are mechanically ventilated in critical care units, with signs and symptoms, such as high body temperature and high numbers of white blood cells (leukocytosis) that are shared by many other disorders critically ill patients may have. Hence, diagnosing, and therefore also treating, VAP is difficult. This has implications for the management of patients with VAP, as it is quite unlikely that Intensive Care Unit (ICU) doctors are able to identify this disorder reliably, and prescribe optimal treatment. It is therefore believed my many clinical experts in the field of VAP that ICU doctors need some form of decision support.

In this paper we describe our work in the design of an architecture for a decision-support system (DSS) that allows evaluating the efficacy of use of the system in a critical-care environment. It is argued that in developing a DSS for clinical use, more attention should be given to evaluation issues than currently is the case. However, the wish to evaluate a DSS has also consequences for its user interface and organisation. In this paper we relate general considerations found in the literature on evaluation to an actual architecture of a DSS.

The structure of the paper is as follows. In the following section, previous research in computer-based clinical decision support is described. In Section 3, the problem of the management of VAP is described and the model that underlies our DSS is summarised in Section 4. In Section 5, we analyse considerations with respect to evaluation for the architecture of the DSS. The resulting architecture is subsequently described in Section 6. The paper is rounded off by some conclusions.

1. Previous Research

Over the last decades, much progress has been made in diagnosing and treating disorders in patients. The medical community itself has introduced clinical guidelines as a vehicle of quality improvement and control, and an increasing number of clinical guidelines have become available. Modern clinical guidelines are evidence-based, i.e. based on results reported in the scientific literature, and are meant to support clinicians in their decision-making process. These clinical guidelines may contain cost components, organisational aspects and aspects for implementation. The main purpose of a clinical guideline is to provide a standard that allows for all physicians to generalise the management of patients with a specific disorder or illness and to reduce variation between physicians [2]. However, very few physicians use a guideline when it is available on paper only. Computerising these clinical guidelines eases their use, but most guidelines are designed with use on paper in mind, and it is therefore not straightforward to convert them into executable form.

An alternative to quality improvement of clinical care is being offered by DSSs. In contrast to medical guidelines, medical DSSs are normally computer-based from the start and thinking about the design of such systems involves taking into account how computer-based systems can best offer support to the clinical user. These computer systems contain medical knowledge provided by medical specialists, but can, of course, also be evidence-based as modern guidelines. By helping physicians in their management process, a decision-support system aims to enhance patient outcomes.

There are many ways in which a DSS can be constructed. As the management of patients with a disorder involves uncertainty, due to the fact that the state of the patient is incompletely known, and tradeoffs between pros and cons of various treatment modalities, we think that Bayesian networks combined with decision theory offers the most suitable basis for such systems [3].

2. The Problem: Incorporating Evaluation into the Design of a DSS

Most medical DSSs are designed and developed taking only into account their final use. However, it has been stated again and again that for DSS to be accepted by the clinical community it is essential that they are properly evaluated. This will have implications for the functionality and organisation of a DSS. For example, to obtain faithful evaluation results it is necessary that a DSS does not influence the decisions made by a clinician. Clearly, this requirement contradicts the final use of a medical DSS, which is expected to support clinical decision making, and thus to improve and change the clinician's decisions. The question, therefore, arises to what extent evaluation requirements are compatible with the intension to produce a final working system, and to which extent incorporated evaluation facilities restrict use of the final system. These questions are being addressed in this paper for the specific system we developed for the management of VAP.

3. A DSS for VAP: its Underlying Bayesian-Network and Decision-Theoretic Model

The DSS for the management of VAP we developed is based on a Bayesian network and also incorporates a decision-theoretic part to make treatment choices. We will only provide a brief summary here; a more detailed description of the Bayesian network and decision-theoretic model, including a motivation for their structure and content, is given in Ref. [7].

A Bayesian network consists of two parts: a qualitative part, i.e. the structure of the network including all relations between the variables, and a quantitative part, i.e. conditional probabilities P(X | pa(X)) for each variable X associated with a vertex V, where pa(X) stands for the set of variables associated with the parents of V. The model for the management of VAP consists of a diagnostic part and a prognostic part. The structure of the diagnostic part is concerned with the diagnosis of VAP, based on the patient's clinical signs, the duration of stay in the hospital and whether or not the patient is mechanically ventilated. To be able to identify the bacteria, which may have been the cause of the pneumonia, the process of bacterial colonisation as taking place in the ICU has also been modelled. For the construction of the quantitative part, the infectious-disease expert had to estimate all conditional probabilities. The prognostic part of the Bayesian network is intended to provide information about the most effective combination of antibiotics, i.e. it is used to determine optimal coverage for VAP. As determining the optimal treatment uses decision theory; the VAP expert was asked to provide utilities for all sensible combinations of antibiotics, taking into account presence and absence of VAP, side effects, financial costs and antimicrobial spectrum.

The global structure of the resulting Bayesian-network model is shown in Figure 1.

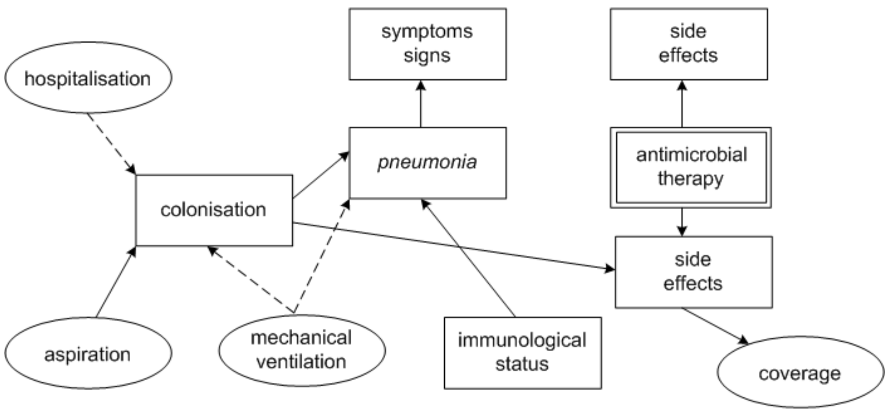

Figure 1. Global structure of the Bayesian-network model for VAP; single-lined boxes stand for sets of variables, and antimicrobial therapy is the decision variable.

4. Design of an Evaluation Study of the DSS for VAP

There are a number of issues that must be taken into account when evaluating a DSS. These issues are described here. We also describe how a resulting system can be evaluated. Subsequently in Section 6 the consequences for the architecture and functionality of the DSS are discussed in detail.

4.1. Evaluation Biases

When using a DSS in a particular user environment, such as the management of a disorder in an ICU, evaluation results may not faithfully reflect the reality due to unknown systematic effects. These confounding effects, or biases, have to be taken into account before designing an evaluation study. Known types of biases include [1]:

- The *volunteer effect*: doctors who volunteer to use the system perform better than others. A possible explanation for this effect is that a volunteer is probably more willing to use the system than someone who is not volunteering;
- The *assessment bias*: knowledge of gold standard or system output may influence the user's decisions. Since there is no real gold standard for the management of VAP, this does not apply. However, when taken into account the system's output, the judgement of a doctor might be influenced by it.
- The *Hawthorne effect*: clinical performance improves if clinicians know they are being studied.
- The *checklist effect*: performance improves if clinicians use a checklist. When using a checklist, a clinician is reminded of the factors that are considered relevant in, for example, diagnosing a certain disorder. This may cause improvement in clinical performance.

Clearly, when evaluating a DSS these biases cannot be ignored, but should be addressed somehow. We will do so in the following.

4.2. Design of the Study

Considering the discussed evaluation biases, we designed the evaluation study for the system as follows. As the Hawthorn effect is difficult to avoid, we will ask *all* ICU doctors to use the decision-support system. By doing so, we make the assumption that all doctors behave like the employees in the Hawthorne factory [8]. Thus, in case usage of our DSS causes a change in diagnostic behaviour, we in a way take into account this bias, as it has been incorporated in all performance measurements. We eliminate the volunteer effect: in this study, as every ICU doctor is expected to use the system when is not particularly reluctant to use the system. As the three ICUs of the University Medical Centre Utrecht that act as our study environment use an electronic patients record system, the doctors are already used to the deployment of a computer-based system in the patient-care process. By making sure that the system does not give an advice concerning a patient before having asked the doctor to enter his or her expert opinion, the possibility that the doctor's original opinion is influenced by the system's output is eliminated.

5. Resulting DSS for VAP

The decision-support system that is described in this paper is intended to be operating within the ICUs of the University Medical Centre Utrecht. Building the system, we had to take the evaluation biases, described in the previous section, into account. As such, it is necessary that the system is integrated with the clinical information system that is used in the ICUs. This system is a full-fletched system that functions as an electronic patient records system (EPR). As no paper patient records are used any more within the three ICUs deploying the EPR, this system contains a wealth of patient information which can be exploited for decision-support purposes.

5.1. Components of the System

The front-end of the DSS consists of a graphical user interface (GUI) with pop-up menus, tables and some graphics, allowing clinicians to enter patient data and to inspect patient data, mostly in textual form, but sometimes – for certain laboratory data – in graphical form, as a time plot. The back-end is a relational database management system, which is linked to the EPR system's relational database system. This not only offers modern facility for secure data storage, updating and retrieval, but also a Structured Query Language (SQL) interface, allowing external systems access to the data.

It was decided to develop a separate user-interface for the decision-support system in particular because this makes separate development possible. We used modern distributed information system technology, in our case Hypertext Preprocessor (PHP). PHP is a server-side HTML-embedded cross-platform scripting language, which allows one to generate content of HTML pages dynamically.

PHP is actually used as the language to link various parts of the system together. The majority of the functionality of the system is offered by a commercially available Bayesian-network and decision-network package, which is used in our project to implement the inference engine for the decision-support system. This part of the system processes patient data, and offers various types of advice based on the results computed after instantiating the Bayesian network model of VAP with the patient data. This advice is presented to the user.

The final components for the system are an HTTP server, and a Web browser. The HTTP server that is used in the project is Apache. The Web browser acts as the user interface to the decision-support system.

The architecture of the system is visualised in Figure 2.

5.2. Added Facilities for Evaluation

For every patient who is mechanically ventilated, the doctor is asked for his or her expert opinion. When a VAP is concerned, also the prescribed antibiotic or combination of antibi-

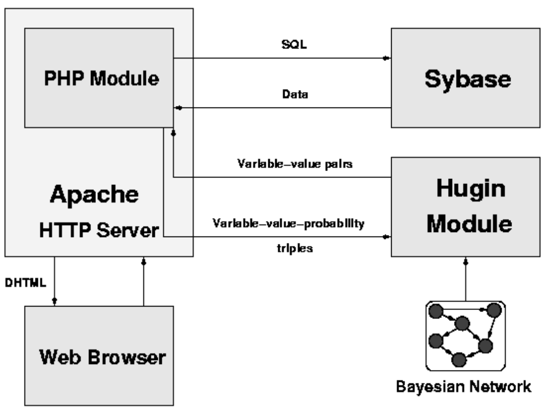

Figure 2. Global architecture of the system.

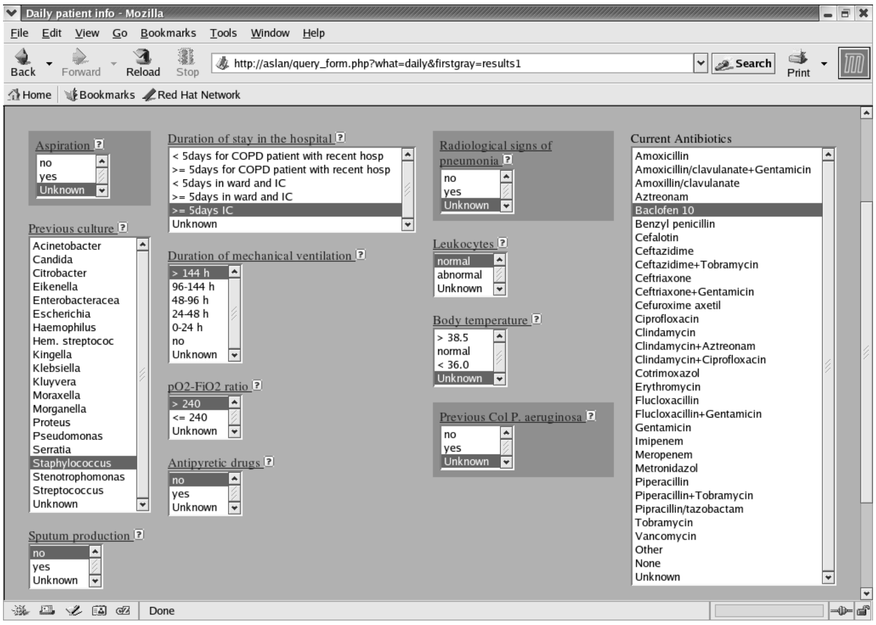

Figure 3. A screenshot of a form that is part of the DSS representing symptoms of the patient.

otics is filled in, including the motivation for this choice. Following this, a form is shown, representing the patient's clinical details. See Figure 3 for a screenshot of this form. These values are pre-selected from the clinical database and, when the doctor disagrees, he or she can change them. This form, or checklist, might be a confounding factor. Unfortunately, measuring or expressing this effect is difficult, if not impossible. But since all physicians are expected to fill in the checklist, measurement of the effect is not necessary. After the physician has filled in the checklist, a management advice is presented to the user *at random*, i.e. on average in 50% of the consultations of the system an advice is obtained and in 50% of cases, no advice is being offered. The advice, if being offered, includes the likelihood for the patient of having VAP and the optimal treatment for possible colonisation of the respiratory tract. By doing so, we are able to compare the two groups of doctors who were or were not presented with an advice. Only when the system's advice is presented, the doctor is asked again whether he or she suspects the patient having pneumonia, and which therapy, i.e. combination of antibiotics, he or she wishes to prescribe, taking into account the system's advice. In this way, we are able to see if the physician changed his or her judgement after the presented advice. One possibility is that the physician has become more aware of some important symptoms of the patient, so that he or she revised the earlier judgement. Another possibility is that the physician did not change his or her decision after the system's output, simply due to ignoring the advice or because filling in the symptoms did not change the physician's view regarding the patient. From this we can draw conclusions concerning the influence of the system's advice on the physician's judgement.

6. Conclusions

Medical decision-support systems are meant to assist clinicians in the difficult process of medical management. As was argued, the availability of medical decision support is even more crucial in a critical care environment, as mistakes made by doctors usually have drastic consequences for the patient. However, certainly in critical care environments doctors are reluctant to use DSS that have not been evaluated clinically. We have, therefore, taking an actual DSS that supports the management of VAP as an example, developed a medical DSS environment that incorporates facilities for clinical evaluation. The reasons for including these facilities were motivated. We believe that more attention should be given to evaluation issues, when developing medical DSSs than has been done so far. When these evaluation issues are incorporated into the design of a decision-support system, it is possible to perform a reliable evaluation of such system by its users even when using it in a real-life setting.

References

[1] Friedman, C. & Wyatt, J. (1997). Evaluation methods in medical informatics. Springer-Verlag, New York. ("Computers & Medicine" series.) ISBN 0-387-94228-9.

[2] Kocher MS, Mandiga R, Murphy JM, Goldmann D, Harper M, Sundel R, Ecklund K, Kasser JR. A clinical practice guideline for treatment of septic arthritis in children. *The Journal of Bone and Joint Surgery (American)* 2003; 85: 994–999.

[3] Lucas PJF, Boot H, Taal BG. Computer-based decision support in the management of primary gastric non-Hodgkin lymphoma. *Methods of Information in Medicine*, 37:206 – 219, 1998.

[4] Bonten MJM, Kollef MH, Hall JB. Risk factors for ventilator-associated pneumonia: from epidemiology to patient management. *Clinical Infectious Diseases* 2004; 38(8): 1141-9.

[5] Bonten MJM, Bergmans CJJ, Ambergen AW, De Leeuw PW, Van der Geest S, Stobberingh EE, Gaillard CA. Factors for pneumonia and colonization of respiratory tract and stomach in mechanically-ventilated ICU patients. *Am J Respir Crit Care Med* 1996; 154: 1339–1346.

[6] Lucas PJF, Van der Gaag LC. *Principles of Expert Systems*. Wokingham: Addison-Wesley, 1991.

[7] Lucas PJF, de Bruijn NC, Schurink CAM, Hoepelman IM. A Probabilistic and decision-theoretic approach to the management of infectious disease at the ICU. *Artificial Intelligence in Medicine* 2000; 19(3): 251–279.

[8] Mayo E. (1933) *The human problems of an industrial civilization* (New York: MacMillan) ch.3.

[9] Berner ES, Maisiak RS, Cobbs GC, Taunton DO. Effects of a Decision Support System on Physicians' Diagnostic Perfomance. *Journal of the American Medical Informatics Association* 1999; 6: 420–427.

Contact information:

Drs. Stefan Visscher (S.Visscher@azu.nl)
University Medical Centre Utrecht
Department of Internal Medicine and Infectious Diseases
Huispostnr. F.02.126
Heidelberglaan 100
3584CX Utrecht